



OPEN ACCESS

EDITED BY

Maria Grazia Sindoni,
University of Messina, Italy

REVIEWED BY

Isabella Poggi,
Roma Tre University, Italy
Andy Lücking,
Goethe University Frankfurt, Germany

*CORRESPONDENCE

Simona Sbranna
✉ s.sbranna@outlook.com

RECEIVED 07 November 2024

ACCEPTED 04 April 2025

PUBLISHED 16 May 2025

CITATION

Sbranna S, Savino M, Baills F and
Grice M (2025) Gaze behaviour and vocal
feedback in task-based dyadic conversations
with and without eye contact.
Front. Commun. 10:1524453.
doi: 10.3389/fcomm.2025.1524453

COPYRIGHT

© 2025 Sbranna, Savino, Baills and Grice. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Gaze behaviour and vocal feedback in task-based dyadic conversations with and without eye contact

Simona Sbranna^{1*}, Michelina Savino², Florence Baills³ and Martine Grice¹

¹IfL-Phonetics, University of Cologne, Cologne, Germany, ²Department of Education, Psychology, Communication, University of Bari Aldo Moro, Bari, Italy, ³Faculty of Arts, Universitat de Lleida, Lleida, Spain

Human communication is multimodal, with verbal and non-verbal cues such as eye gaze and vocal feedback being crucial for managing interactions. While much research has focused on eye gaze and turn alternation, few studies explore its relationship with turn-regulating vocal feedback. This study investigates this interplay during a Tangram game in Italian under two visibility conditions: face-to-face and separated by a screen. The results show that feedback producers rarely look at receivers, while receivers more frequently look at producers, suggesting that they might be eliciting vocal feedback. Without visual contact, gaze shifts decrease and vocal feedback increases. Interestingly, when visual contact is absent, gaze directed towards where the addressee is sitting does not coincide with vocal feedback, raising questions about what prompts this gaze.

KEYWORDS

turn-taking, eye gaze, vocal feedback, face-to-face dialogue, non-visibility

1 Introduction

Human communication integrates several information channels, such as words, voice modulation (i.e., prosody), facial expressions, and gestures, enhancing the overall communicative impact in conversations (e.g., [Paulmann et al., 2009](#)). This is particularly true in face-to-face interactions, where all these elements play a role in regulating social interaction and help to jointly create meaning (e.g., [Rasenberg et al., 2022](#)). During conversation, interlocutors maintain the flow of their dialogue by speaking one at a time and taking turns in a cooperative manner. Interactions have been described as complex dynamic systems ([Hollenstein, 2013](#); [Cameron and Larsen-Freeman, 2007](#)) since they emerge dynamically and evolve through a process of co-regulation between interlocutors ([Hu and Chen, 2021](#) for a review; [Stahl, 2016](#) for task-based dialogues). Turn-taking is believed to be driven by social signals consisting of verbal or non-verbal cues exchanged between individuals engaged in a conversation (e.g., [Kendrick et al., 2023](#)). These signals serve as indicators of readiness to shift from speaking to listening or willingness to take the turn in speaking. Gaze cues (i.e., the way speakers orient their gaze) and vocal feedback (i.e., signals of active listening) have been shown to play important roles in the management of turn-taking in conversations. However, in relation to speakers' intention of taking or not taking the floor, they have been traditionally investigated as separate entities (e.g., see [Degutye and Astell, 2021](#) for a review of studies on eye gaze and turn alternation, and [Drummond and Hopper, 1993](#) as an example of a study on vocal feedback and turn alternation).

In the present study, we explore the relationship between turn-regulating vocal feedback and eye gaze behaviour in dyadic task-based conversations to shed light on these two

mechanisms as part of multimodal communication. Specifically, we examine gaze behaviour (direction, duration up to the production of vocal feedback, and possible shift afterwards) in both the feedback producer (the listener) and feedback receiver (the speaker) when two types of turn-regulating vocal feedback are produced. These two types of feedback refer to feedback when the feedback producer does or does not subsequently take the floor. We additionally compare results across two visibility conditions: when the visual channel is available (with eye contact) and when it is not available (without eye contact). The aim of the latter manipulation is to investigate how far the overall gaze and feedback behaviour, as well as the relationship between them, changes when only the auditory channel is available.

Below, we review research on eye gaze and turn alternation in Section (2.1) and vocal feedback in Section (2.2) and discuss the interplay of vocal and visual cues for turn alternation in Section (2.3). Since we are concerned with the difference between dialogue with and without eye contact, we review research on the impact of visibility of the interlocutor in Section (2.4). In Section (3), we list and discuss our research questions; in Section (4), we provide information on the materials and methods; in Section (5), we present the results; and finally, in Sections (6) and (7), we discuss our findings in relation to previous findings, as well as limitations and future directions, respectively.

2 Background

2.1 Eye gaze and turn alternation

When individuals engage in social interactions, these exchanges are naturally organised into distinct segments known as turns, representing the time during which one person speaks before the other takes the floor. Early studies have looked at conversational partners' gaze in relation to their alternating roles as speaker and listener (as in Goodwin, 1981, which we also refer to as primary and secondary speakers, respectively, since the listener in fact speaks when giving vocal feedback). These studies claim that in dyadic interactions, people tend to look at the other participant more when they are listening than when they are speaking (Argyle and Cook, 1976, among many others). Kendon (1967, 1990) provided a more precise description of the different patterns of speaker and listener gaze: Listeners tend to maintain long gazes at speakers, interrupted by brief glances away, while speakers alternate gazes towards and away from the listener with approximately equal gaze durations.

In addition to these general gaze patterns, researchers have looked more specifically at what happens at crucial conversational moments such as turn transitions. In speech, turns have generally been found to transition from one speaker to another with brief gaps and few overlaps (Levinson and Torreira, 2015). This pattern, which appears to be highly robust in the face of individual, methodological, and contextual variation, raises interest in the nature of the factors that regulate this remarkable synchronisation between interlocutors.

The multimodal nature of turn transitions is widely attested (see, for example, Kendrick et al., 2023 for the effect of manual and gaze signals on turn transitions), as is the role of eye gaze in facilitating turn-taking (see Degutye and Astell, 2021 for an extensive review). This was established in the seminal work of Kendon (1967), who has been highly influential in understanding the role of eye gaze at the

start and end of a speaker's turn. In a corpus of spontaneous conversations, he found that over 70% of utterances began with the speaker looking away from the listener, while over 70% ended with the speaker looking at the listener. Kendon argued that at the beginning of their turn, participants in a conversation tend to avert their gaze (i.e., they look away from their interlocutor). This serves two functions: first, indicating that they are focusing internally, concentrating on formulating their thoughts, planning their speech, or recalling information before they begin to speak; and second, functioning as a signal to others that they intend to take the floor by initiating a speaking turn. Thus, by averting gaze, the participant who currently holds the conversational floor signals that they are not available as a *listener*.

These observations were confirmed by a number of subsequent studies on dyadic exchanges (e.g., Cummins, 2012; Duncan, 1972; Duncan and Fiske, 1977; Ho et al., 2015; Oertel et al., 2012; see also Jokinen et al., 2009 for natural three-party conversations). For instance, Novick et al. (1996) found a systematic temporal alignment between gaze and speech during turn-taking. In that study, dyads were recorded while playing guessing games using an eye tracking device. They found that at a turn transition, the speaker ends their turn looking at the listener, and the listener begins to speak with an averted gaze. Although Kendon's observations have been confirmed by the aforementioned studies, a number of studies have attributed gaze behaviour during turn-taking to different factors (Beattie, 1978; Rossano, 2012; Rutter et al., 1978; Streeck, 2014). Beattie (1978) argued that the speaker's gaze away during early utterance production (and during re-engagement at the end of the utterance) is solely driven by the necessity to reduce cognitive load and does not have any regulatory role during turn-taking. Alternatively, Rossano et al. (2009) and Streeck (2014) claimed that gaze does not facilitate turn-taking as such, but it does facilitate the organisation of complex actions that might require multiple turns to complete.

At the end of a turn, Kendon (1967) proposed that participants gaze towards the listener with the function of checking their availability as the next speaker and, thus, as a signal of turn-yielding. This proposal is strongly supported in later studies (Novick et al., 1996; Rutter et al., 1978; Lerner, 2003; Jokinen et al., 2009, 2013; Ho et al., 2015; Br ne et al., 2017; Auer, 2018; Blythe et al., 2018; Streeck, 2014; Oertel et al., 2012; Kawahara et al., 2012; for group conversations see also Harrigan and Steffen, 1983; Kalma, 1992). However, Kendon's findings were not supported across the board in this context either. Rutter et al. (1978) commented that Kendon's predictions about floor changes can only occur if the listener is also looking at the speaker, that is, in the case of mutual gaze between the two participants. Moreover, Novick et al. (1996) observed two prevalent patterns of gaze direction during turn-taking, categorised as 'mutual-break' and 'mutual-hold'. At the end of an utterance, the primary speaker (the participant currently holding the floor) looks at the secondary speaker (the participant currently listening), at which point the gaze is momentarily mutual until the secondary speaker breaks the mutual gaze and starts to speak, hence 'mutual-break', which is the most frequent pattern. Alternatively, in instances of 'mutual-hold', mutual gaze is held while the secondary speaker starts speaking without immediately averting their gaze. Beattie (1978), in fact, found the converse to Kendon (1967), with more immediate speaker switches when utterances terminated with no gaze than with gaze, and no differences in the effect of gaze on the number of immediate speaker

switches for complete utterances, as might be expected if gaze served a floor-allocation function.

Degutye and Astell (2021) argued that different results relative to turn boundaries and gaze might be due to different experimental designs and different types of interaction (e.g., dyadic vs. triadic/multiparty interaction, question/answer sequences vs. other types of adjacency pairs in free conversation) and methodology (see also considerations of the same kind in Spaniol et al., 2023). In addition, many individual factors can affect gaze behaviour during conversation (gender in Argyle and Dean, 1965; Myszka, 1975; Bissonnette, 1993; social status in Myszka, 1975; Foulsham et al., 2010; acquaintance status in Rutter et al., 1978; Bissonnette, 1993; and cultural background in Rossano et al., 2009).

While there is extensive research on eye gaze during turn alternations, there is less research on how eye gaze interacts with other aspects of turn alternation. One such aspect is turn-regulating vocal feedback, which is discussed in the next section.

2.2 Vocal feedback and turn-regulating functions

The speaker's planning of their contribution to a conversation is significantly influenced by a crucial variable: the listener's reaction to their utterance, realised through vocal feedback signals. While often overlooked in conversation (Shelley and Gonzalez, 2013), these vocal feedback signals play a pivotal role in guiding verbal exchange, indicating the listener's attitude towards what is being said and conveying intentions related to floor management. Commonly referred to as 'backchannels' or 'response tokens', these signals are argued to enhance fluency in social interactions by supporting the ongoing turn of the interlocutor (Amador-Moreno et al., 2013) and structuring dyadic conversations (Kraut et al., 1982; Sacks et al., 1974; Schegloff, 1982).

There is no consensus in the literature about their definition (Rühlemann, 2007). Fries (1952) is credited as one of the earliest to identify these 'signals of attention' which do not disrupt the speaker's discourse in telephone conversations. Over time, various terms have been employed to characterise this phenomenon, including 'accompaniment signals' (Kendon, 1967), 'receipt tokens' (Heritage, 1984), 'minimal responses' (Fellgely, 1995), 'reactive tokens' (Clancy et al., 1996), 'response tokens' (Gardner, 2001), 'engaged listenership' (Lambertz, 2011), and 'active listening responses' (Simon, 2018). Yngve (1970) introduced the term 'backchannel communication' to distinguish the primary channel used by the speaker holding the floor from the one used by the listener to convey essential information without actively taking a turn. Essentially, backchannels represent tokens used to express acknowledgement and understanding, encouraging the main speaker to continue (e.g., Hasegawa, 2014, among others).

Later, Jefferson (1983) claimed that the use of such 'acknowledgement tokens' does not exclusively signal turn-yielding but can also cue the intention of taking the floor after acknowledging the current speaker's turn. In particular, she observed that some token types "exhibit a preparedness to shift from reciprocity to speakership" (1983:4), whereas others are more systematically used as turn-yielding signals. For English, she noticed that speakers tend to use tokens such as 'mh-mh' to signal the intention of acknowledging without taking

the floor, whereas tokens such as 'yeah' when a recipient is moving into speakership. In line with this distinction, some later studies (Drummond and Hopper, 1993; Jurafsky et al., 1998; Savino, 2010, 2011, 2012; Savino and Refice, 2013; Wehrle, 2023; Janz, 2022; Sbranna et al., 2022; Spaniol et al., 2023; Sbranna et al., 2024) categorised backchannels based on two turn-taking functions, that is, passive reciprocity (henceforth PR) and incipient speakership (henceforth IS). Passive reciprocity involves tokens produced without the speaker taking the floor, serving as acknowledgements and continuers (i.e., backchannels, as intended by Yngve, 1970). Incipient speakership refers to tokens used by a speaker to acknowledge the interlocutor's turn before producing a new turn themselves, realising a turn transition. This classification has also been partially supported by intonation research, which has observed a tendency for certain tokens and intonation contours to co-occur with these functions (Savino, 2010, 2011; Sbranna et al., 2022; Sbranna et al., 2024).

In this study, we adopt this operationalisation of the feedback turn-taking function, as it accounts for the role that vocal feedback signals play in the complex and multidimensional turn management system, alongside feedback signals of a different (i.e., visual) nature.

2.3 The interplay of vocal and visual cues for turn alternation

Given that feedback is meant to support the ongoing turn of the primary speaker, this speaker can actually invite the secondary speaker to produce feedback. This invitation is multimodal in nature and can be signalled in different ways, for example, by inserting a pause by intonation, or head or gaze movements. Heldner et al. (2013) proposed that the primary speaker provides "backchannel relevance spaces" to allow the listener to insert backchannels, although in their corpus of face-to-face interaction, the number of such spaces was considerably greater than the actual backchannels provided. The authors argue that not all backchannel relevance spaces are filled (either with vocal or visual feedback) as listeners choose feedback positions that actively support speakers in constructing their discourse.

Some studies on the interaction between vocal feedback and, in particular, eye gaze have shown that vocal and non-vocal feedback signals are used to show acknowledgement and understanding of the primary speaker without taking the floor—those which we refer to as PR tokens, and Heldner's 'backchannels' (2013)—occur during mutual gaze between a listener and a speaker, in accordance with a number of other studies (Kendon, 1967; Bavelas et al., 2002; Eberhard and Nicholson, 2010; Cummins, 2012; Oertel et al., 2012). For example, Oertel et al. (2012) compared gaze behaviour in dyadic interactions at turn transitions during speech overlap, in silence, and in the vicinity of backchannels and found that the production of vocal feedback signals is associated with an increase in mutual gaze. At a more detailed level, Bavelas et al. (2002) found that gaze patterns used to coordinate feedback signals in dialogue are often initiated by the speaker gazing at the listener, who then looks back, resulting in short periods of mutual gaze broken by the listener looking away shortly afterwards. The authors precisely examined when and how listeners insert their feedback into a speaker's narrative by involving participants in a storytelling task. A collaborative approach would predict a relationship between the speaker's acts and the listener's responses, and the authors proposed that gaze coordinates this

collaboration. They found that the listener typically looks at the speaker more often than the other way around. However, at key points in their speech, the speaker seeks a response by looking at the listener, creating a brief period of mutual gaze, called “gaze window” by the authors. They observed that, during gaze windows, the listener was very likely to provide a reaction, such as a vocal ‘mhm’ or a nod, after which the speaker interrupted the gaze window by quickly looking away and continuing to speak. This pattern aligns with previous studies on the use of multimodal cues in the perception of interrogativity, which show that gaze direction towards the interlocutor and other gestural signals enhance the perception of a response being expected (Borràs-Comes et al., 2014).

Although mutual gaze has been found to be a strong predictor of feedback signals independent of their modality (Ferré and Renaudier, 2017; Hjalmarsson and Oertel, 2012; Poppe et al., 2011), a difference has been reported between vocal and non-vocal (i.e., head gestures) feedback signals, with non-vocal signals being more often produced in correspondence with mutual gaze than vocal signals (Ferré and Renaudier, 2017; Eberhard and Nicholson, 2010; Truong et al., 2011). This is probably because vocal signals do not need to be conveyed through the visual channel. Moreover, gaze was found to be sustained more often throughout a sequence that contained a visual backchannel than in a sequence containing a verbal backchannel (Ferré and Renaudier, 2017). The strong correlation between being looked at and gestural backchannels, but not vocal ones (Bertrand et al., 2007), was motivated by the idea that gaze might establish a communication mode between interlocutors. Nonetheless, in contexts where the primary speaker was not gazing at the interlocutor, both gestural and/or vocal backchannels were produced.

Onďáš et al. (2023) also looked at the temporal detail of the interval between the start of the gaze directed towards the interlocutor and the subsequent backchannel in an interview scenario and found that, in most cases, the moderator received vocal feedback from his guest within 500 ms of initialising direct eye contact. The second most probable time interval was 1,500–2000 ms. This same interval was found to be the most probable in the opposite scenario, that is, the moderator providing feedback to the guest. One possible speculation on this different behaviour is that the moderator is meant to lead the conversation, thus getting quicker reactions from the guest, who is expected to inform the moderator that they acknowledge and follow the dialogue structure. However, this result is linked to a very specific conversational format, which probably has its own dynamics that are different from other formats of conversation. Moreover, Kendrick et al. (2023), analysing the impact of visual cues in turn-timing, found that turn transitions were sped up by the use of manual gestures but not by gaze behaviour.

These studies focused on the interaction between eye gaze and backchannels, accounting only for their PR function. To our knowledge, only one recent exploratory study (Spaniol et al., 2023) has investigated the interplay of gaze patterns and turn-regulating vocal feedback using the same PR vs. IS paradigm for vocal feedback. In particular, the authors analysed gaze and vocal feedback behaviour in dyadic conversations across different communicative contexts, namely, (1) a free conversation in which participants get to know each other, (2) a Tangram game-based interaction, and (3) a free discussion about the game participants had just played together. In contrast to previous findings, they found that in the task-based context, vocal feedback was mostly provided in conjunction with averted gaze,

independently of its turn-regulating function. They explained their results by the presence of a visual competitor during the conversation, that is, the task material. In the other two contexts of free conversation, three out of four dyads produced PR feedback signals more under gaze directed to the interlocutor than IS feedback signals, showing that the turn-regulating function of backchannels distributes differently on gaze behaviour patterns and that this distribution is influenced by the context of the conversation.

2.4 The impact of (non-)visibility on communicative channels

Another important question for the present study concerns the use of speech and gestures when the interlocutor is not visible.

Some studies have investigated the use of manual gestures in conversations, with and without visibility between participants, concluding that gestures have both communicative and internal cognitive functions. For example, Alibali et al. (2001) found that participants were still using representational gestures, that is, gestures depicting semantic content related to speech, when an opaque panel hindered mutual visibility, although to a lesser extent than in visibility conditions. Interestingly, the number of beat gestures, that is, simple rhythmic gestures lacking semantic content, remained consistent across both conditions. Bavelas et al. (2008) compared the use of gesture in face-to-face and telephone conversations and concluded that, instead of being incidental, the gestures made while talking on the phone appeared to be a deliberate adjustment to a dialogue without visual contact. Indeed, both the visibility and non-visibility groups gestured at a high rate, although in the visibility condition, participants gestured with forms and relationships with words that were more informative for their listeners. These studies point to similar conclusions: first, that gestures are adjusted to the interlocutor's need; second, that gestures fulfil not only a communicative function but also an internal cognitive function (see also McNeill, 2005).

Based on these findings, an interesting question is whether these results apply to eye gaze. Argyle et al. (1973) sought to distinguish the various functions of gaze by utilising a one-way screen to manipulate the conditions of speakers observing or being observed. They found evidence of gaze having monitoring and signalling functions. In the first case, they observed that individuals who could see through a one-way screen exhibited increased gaze while speaking compared to those without visual access; in the second case, they observed that even the individuals who could not see their interlocutor still directed their gaze towards the other person occasionally while speaking. Indeed, several studies have found that, even in the absence of visibility, listeners are particularly sensitive to sound sources directed straight at them and are able to recognise the speaker's head orientation, that is, the situation corresponding to mutual gaze in visible conditions (Edlund et al., 2012; Kato et al., 2010; Nakano et al., 2008).

In a study examining mutual gaze patterns in dyadic conversation under light and darkness conditions, Renklint et al. (2012) found a notable reduction in instances of mutual gaze in the absence of light. To interpret this result, the authors claimed that mutual gaze is “made for the other person to see” (aligning with the perspective of Bavelas et al., 1992:483) and that gaze extends beyond mere learned behaviour, having its own functions and meanings. Moreover, the authors noticed that in dyadic conversations, the addressee is consistently the only other

participant in the conversation, and given an implicit mutual agreement about an alternation of speakers in a conversation, selecting the next speaker becomes a mere formality. This is very different from multiparty conversations, in which gaze plays a significant role in turn alternation. Thus, they suggested that in dyadic conversations, information derived from linguistic and phonetic aspects (e.g., sentence structure and prosody) provides enough cues to predict turn alternation, making mutual gaze less essential for successful communication. From these studies, it appears that (i) gaze clearly has a dual function: sensing (i.e., sampling information from a visual scene) and signalling (i.e., not just to gain information from the environment, but with an explicit and deliberate communicative goal) (Argyle and Cook, 1976; Gobel et al., 2015; Risko et al., 2016; Cañigual and Hamilton, 2019), and (ii) in the absence of visibility, gazing at the position where the conversational partner is known to be is reduced but is not absent. This might be due to the fact that in such a condition only, the sensing function of gaze applies.

The effect of non-visibility also relates to the fluency of the speakers' individual speech and the turn alternation system, both of which vocal feedback crucially contributes to. Beattie (1979) reported that the absence of gaze leads to an increased use of filled pauses (see also Cook and Lalljee, 1972) and interruptions (Argyle et al., 1968). Similar results were obtained by Boyle et al. (1994), who studied the effect of visibility on task resolution. The authors found that in the visibility condition, there is greater efficiency in the dialogues attributed to the exchange of visually transmitted non-verbal signals, while in the non-visibility condition, oral communication is characterised as having greater flexibility and versatility. In the latter case, participants tried to compensate for the absence of the visual channel by interrupting their partners more frequently and using more vocal backchannels to support the primary speaker. In addition, Neiberg and Gustafson (2011), comparing eye-contact to non-eye-contact conversations, found an increase in overlapped turn transitions in the absence of visibility, which they explain by a possible increase in unintentional interruptions. From these findings, it seems that visibility improves the smoothness of the dialogue flow, but, in turn, in the absence of visibility, speakers rely more on vocal resources, especially feedback signals.

3 Research questions

Previous research has established the importance of eye gaze, particularly eye gaze direction, for control and coordination in human interactions. However, little is known about the interplay between eye gaze and vocal feedback encouraging the interlocutor to continue speaking (Passive Reciprocity), and feedback initiating a turn (Incipient Speakership). We investigate this relationship in dialogues between Italian speakers playing a Tangram game. Importantly, this game was played in two visibility conditions, with and without eye contact, to compare how turn-taking is regulated with and without gaze being used for this purpose.

In particular, we aim to answer the following research questions:

- 1 How do gaze and vocal turn-regulating feedback work together?
- 2 What is the effect of (non-)visibility of the interlocutor on the feedback-gaze relation?

Our first research question is based on the findings reported above that turn initiation preferably co-occurs with speakers tending to gaze

away from their interlocutor and turn yielding with speakers looking at their interlocutor. Given that IS vocal feedback initiates a turn, we predict behaviour similar to that reported for turn initiation. By analogy, since PR vocal feedback producers do not take the turn from the interlocutor, we expect similar behaviour to that found for turn yielding. We provide our predictions for the feedback producer and receiver separately as follows:

- 1a Feedback producer (listener or secondary speaker):
 - IS vocal feedback, being turn-initiating, is produced while looking away from the feedback receiver to signal unavailability as a listener and commitment to speech planning.
 - PR vocal feedback, which does not involve turn transition, is produced while looking at the primary speaker to show availability as a listener.
- 1b Feedback receiver (primary speaker):
 - If gaze is primarily used to request feedback, vocal feedback should be produced when/just after the primary speaker looks at the secondary speaker.
 - If gaze at the conversational partner is intended to signal turn yielding, the feedback receiver is expected to look away when receiving PR feedback so as not to yield the turn, but to look at the partner when receiving IS feedback, to signal their availability as a listener and thus their willingness to yield the turn.

In the case of the feedback receiver, it is important to keep in mind that the signalling function can be two-fold: feedback request and/or turn holding/yielding, and that the two communicative intentions can overlap, resulting in a more complex picture. Although the feedback receiver is not in control of the actions of the feedback producer, such that speakers may not align in the signals they send with regard to willingness to yield and take the turn, we predict that their behaviour will be somehow cooperative. Based on similar measures previously used in the literature, we explore this cooperativeness through:

- 1c The temporal details of gaze direction intervals (i.e., intervals of time in which participants hold their gaze in a specific direction, delimited by gaze shifts from and to another direction. See Figure 1 in the Method section) that induce vocal feedback to assess whether gaze directed to the addressee speeds up feedback production and/or turn alternation, a topic about which the literature does not provide enough information to build expectations (see Section 2.3).
- 1a The shift in gaze direction after vocal feedback utterance. We expect that the gaze window will close after vocal feedback, as suggested by previous studies. In other words, if gaze towards the addressee is present before the vocal feedback signal, it shifts to the task afterwards.

Relative to our second research question, based on previous findings, we expect that in the non-visibility condition:

- 2a We will not find many co-occurrences of gaze and vocal feedback as multimodal cues for turn regulation, with only the vocal channel being available.
- 2b Gaze towards the addressee will be drastically reduced.

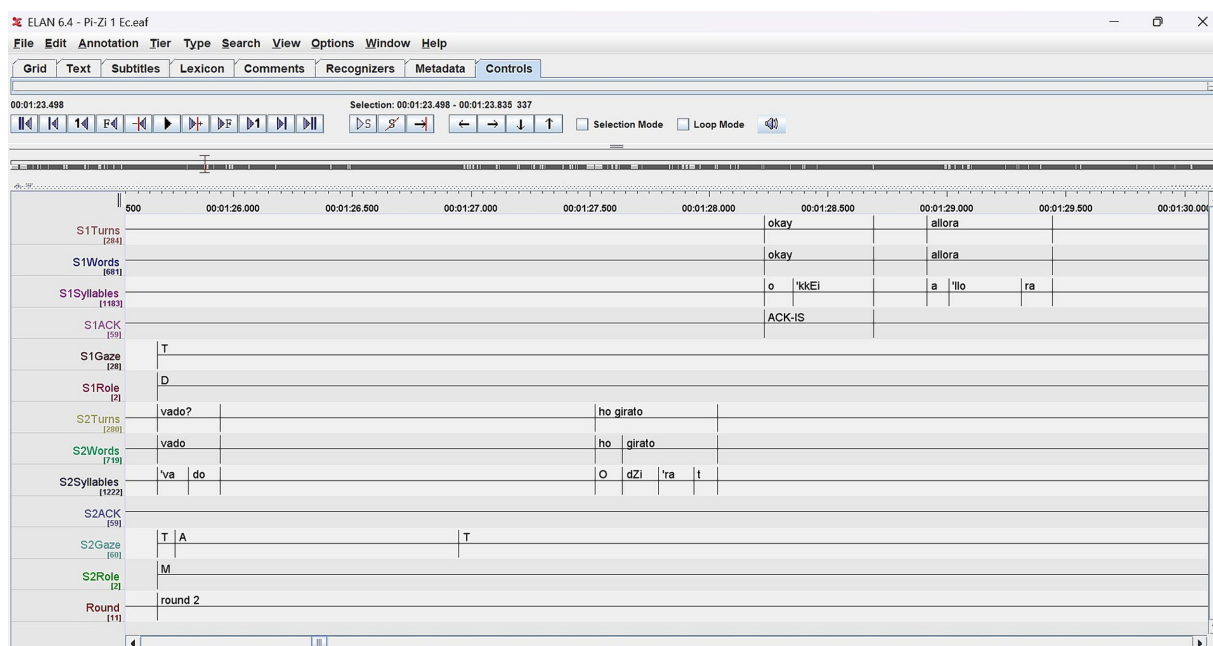


FIGURE 1

Example of all annotation tiers in ELAN. The labels for Role are either “D” for Director or “M” for Matcher. The labels for Gaze are “T” for task-directed gaze, “A” for addressee-oriented gaze, or “O” for other, under which we define gaze to the experimenter. In the tier “ACK” (acknowledgements), the turn-regulating function of the vocal feedback is annotated as IS for Incipient Speakership and PR for Passive Reciprocity.

2c Conversely, the use of vocal feedback will be enhanced to compensate for the lack of visual cues.

These expectations are linked to the assumption that gaze has a dual function of sensing and signalling and that the latter is only possible when mutual gaze is available.¹

4 Methods and materials

4.1 Corpus

We analysed interactions from an Italian corpus of dyadic task-oriented conversations (described in [Savino et al., 2018](#)). As this is the first multimodal analysis of the corpus, the corpus and setup are described in detail below.

4.1.1 Participants

The participants were 12 Italian speakers (6 dyads), all students at the University of Bari, and all from the same geo-linguistic area (the Bari district in Apulia, a southeastern region of Italy). They were all young female adults (aged 21–25 years) and university classmates, ensuring a degree of familiarity within the dyads. Keeping these

factors constant was crucial since gender and acquaintance status have been shown to affect gaze behaviour (see [Myszka, 1975](#) and [Bissonnette, 1993](#), respectively). All speakers voluntarily participated in the experiment and signed an informed consent form. They obtained a course credit for participating in the experiment.

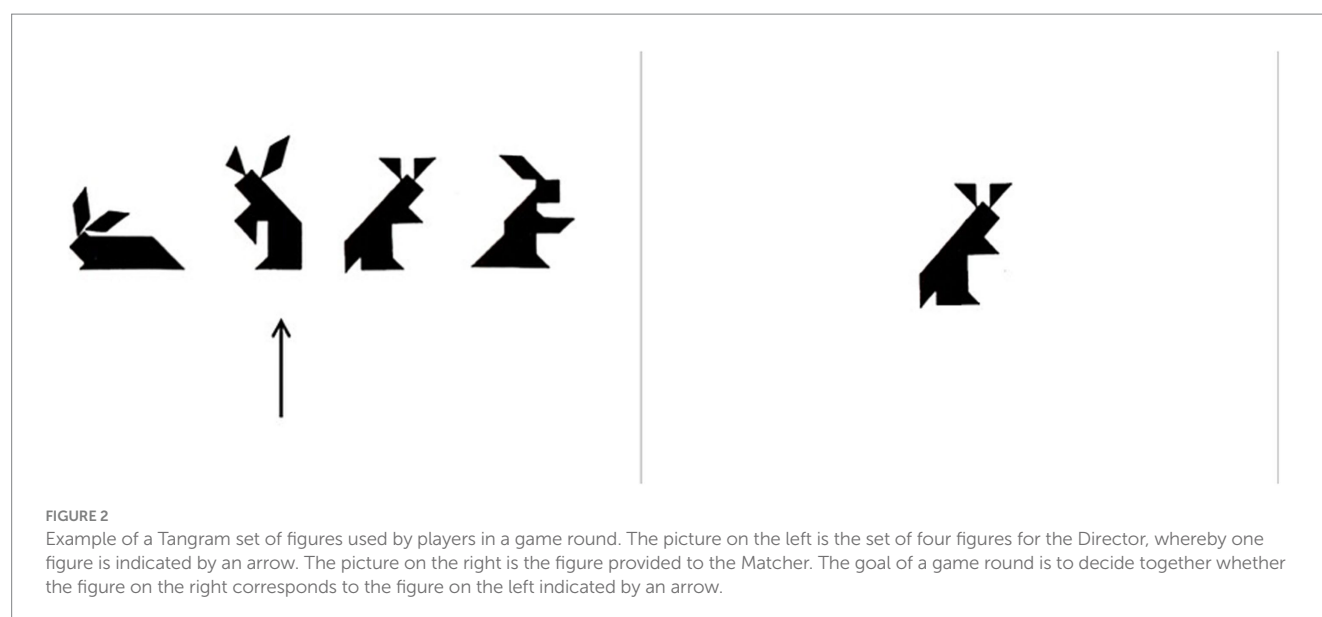
4.1.2 Elicitation method

Participants were asked to play a tangram-based matching game, organised in 22 rounds. For each game round, the two participants were given sets of Tangram figures according to their role in a particular game round: Director or Matcher.² The Director was provided with a sheet with four Tangram figures on it, one of which was marked by an arrow, and the Matcher was given another sheet with only one of the figures belonging to the Director's set (an example of both types of figure is provided in [Figure 2](#)). Participants were unable to see their partner's figure(s), and the goal of each game round was to establish whether the Tangram figure given to the Matcher corresponded to the figure marked by the arrow in the Director's set. A round is defined as each game dialogue segment starting from when participants uncover a set of Tangram figures and finishing when they reach their joint decision as to the matching/mismatching for that set.

In the instructions, the Director was first asked to describe the figure indicated by an arrow, after which the participants could exchange information about the shape of their respective figures.

¹ Please note that by using the opaque panel in the non-visibility condition, we restrict the amount of information that can be sampled from the environment, but not the sensing function itself, as was done in previous studies comparing light vs. darkness conditions ([Renklint et al., 2012](#) as mentioned in Section 2.4).

² Participants alternated their role as Director or Matcher in each round, so that the distribution of role type was balanced between partners across the whole recording session, i.e., each speaker of a dyad played the role of Director 11 times and that of Matcher 11 times.



Information exchange could be freely and spontaneously managed by the participants until an agreement was reached regarding whether the two figures were the same. A typical game round started with the Director describing their Tangram figure (indicated with an arrow) to the Matcher and ends with the two participants checking whether their joint decision about the matching/unmatching figures was correct. The players were explicitly instructed to arrive at a decision based on common agreement. To encourage cooperative behaviour, they were told that they would both score a point every time they correctly identified whether the figure was the same or not and that they would both lose a point when their guess was wrong.

4.1.3 Recording sessions

During the recordings, the two participants sat at separate desks facing each other. A Panasonic HC-V700 camcorder was placed behind each participant for the video recording. The camcorder was placed to clearly capture the participant's visible upper body and face (see Figure 3).

In the eye-contact (henceforth EC) condition, in which the two participants *were able to see* each other, a low opaque panel was placed between the participants' desks at a suitable height to prevent them from seeing each other's Tangram figures while preserving eye contact between players. In the no-eye-contact (henceforth NEC) condition, in which the two participants *were not able to see* each other, a higher opaque panel was placed between the participants' desks so that the conversational partner was completely hidden. The NEC game sessions were recorded 1 month after the EC sessions. The sets of Tangram figures differed across the two sessions.

A total of 12 recordings were obtained (six dyads × two conditions): two recordings for each dyad, one in the EC and one in the NEC condition. Each recording lasted approximately 30 min. All recordings were carried out in a quiet room at the Department of Education, Psychology, and Communication of Bari University, Italy.

In the present study, we analyse a subset of this corpus, consisting of six rounds per dyad in each of the two visibility conditions.

4.1.4 Annotation of vocal feedback expressions

Manual annotation of the speech signal was carried out at various levels in Praat (Boersma and Weenink 2001), including intervals

corresponding to game rounds, inter-pausal units (IPUs, defined as portions of speech delimited by at least 100 ms of silence), phonological words, and syllables.

A specific annotation tier (see Figure 1) is devoted to the description of vocal feedback expressions in relation to their turn-regulating function. In this tier, all lexical tokens, such as *si* (yes), *esatto* (exactly) *va bene* (alright), and *okay*, and non-lexical tokens, such as *mh-mh* and *eh*, signalling attention/understanding/acknowledgement of the current speaker were annotated as feedback signals. To code their turn-regulating functions, we adopted the pragmatic distinction between PR and IS described in the literature review (Section 2.2), resulting in the following operational criteria:

- 1 When the secondary speaker did not take the floor after producing the vocal feedback (so that the primary speaker continued talking), that is, the vocal feedback signal was *not followed* by a turn alternation, the feedback expression was coded as fulfilling a PR function.
- 2 When the secondary speaker took the floor after producing the vocal feedback, that is, the feedback production was *immediately followed* by a turn alternation, that feedback expression was coded as fulfilling an IS function.

These labelling criteria are exemplified in the following example extracted from the corpus:

Director: <ehm> sulla destra c'è <ee> un triangolo rettangolo.

(Eng.: <um> on the right there is <er> a right-angled triangle)

Matcher: <m> [vocal feedback-PR].

Director: cioè su <ehm> mentre la base è formata da un parallelepipedo a sinistra e un triangolo con la punta rivolta verso il basso.

(Eng.: that is on <uhm> while the base is formed by a parallelepiped to the left and a triangle with the tip of the triangle facing downwards)

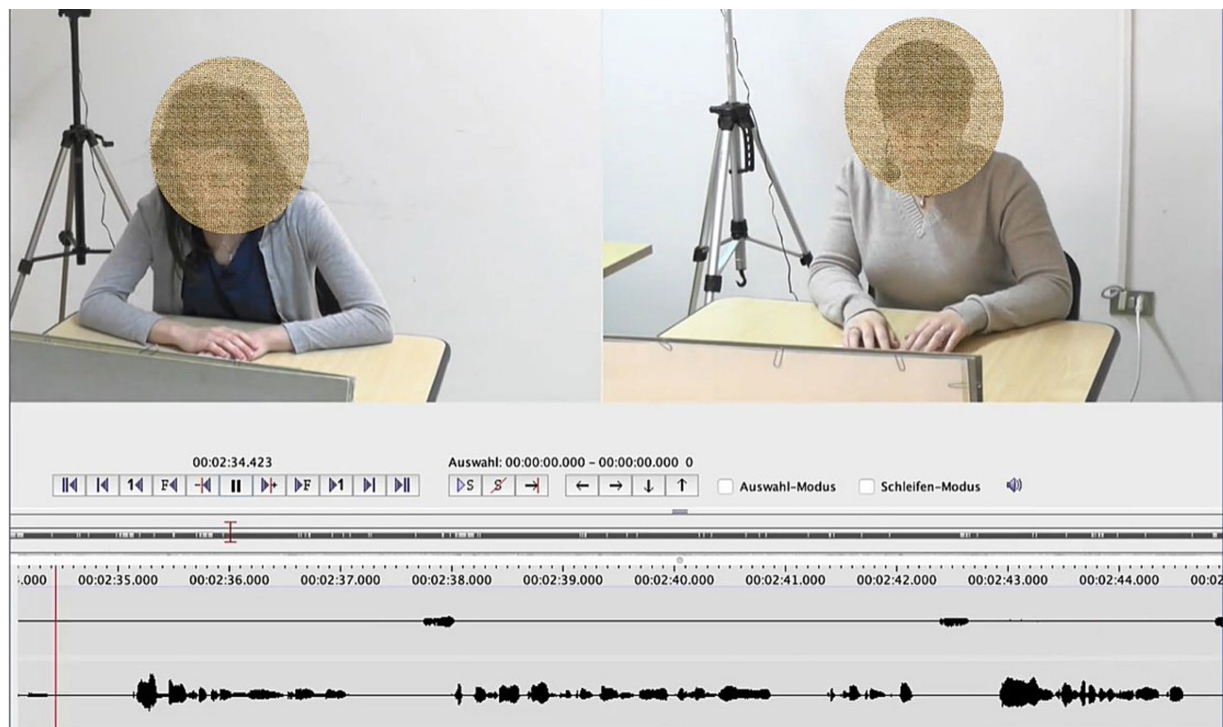


FIGURE 3

Example of the recording setup from the perspective of the two cameras in the Eye-Contact condition.

Matcher: okay [vocal feedback-IS] la base ci siamo descrivimi la vela.

(Eng.: okay for the base we are there, describe the sail)

4.1.5 Annotation of eye gaze direction intervals

The annotation of gaze was performed using ELAN (The Language Archive, 2023; Wittenburg et al., 2006). Following the methodology adopted in previous studies (Kendon, 1967; Beattie, 1978; Goodwin, 1980; Egbert, 1996; Novick et al., 1996; Jokinen et al., 2009; Streeck, 2014; Auer, 2018; Blythe et al., 2018), we annotated intervals of specific gaze direction, namely, the time intervals in which participants gaze continuously in a specific direction without a perceivable change, and where gaze boundaries are identified by a gaze shift from and to a different direction. Gaze direction intervals were equally defined in the EC and NEC conditions for ease of comparison: “Task,” “Addressee,” or “Other.”

In the EC condition, gaze to “Task” was annotated when the participant gazed at the table, where their own sheet with the Tangram figures was positioned. Gaze to “Addressee” was annotated when the participant was looking straight ahead at the other participant. Finally, gaze to “Other” includes gaze directed towards the experimenter.

In the NEC condition, gaze to “Task” was annotated following the same criteria as in the eye-contact condition, whereas gaze to “Addressee” was interpreted as *addressee-oriented* gaze. To annotate it, our main criterion was a head movement upward towards the panel behind which the interlocutor was sitting, accompanied by gaze that could be directed straight ahead, to the top-right or top-left, but without a change in the head position, which remained straight,

directed towards the panel. A change in pupil direction was not interpreted as attention towards something else, since it appeared that participants were focused on their interlocutor and the conversation, and that changes in pupil direction were not related to listening or speaking, as the top-right or top-left pupil movements co-occurred with both speech and silence. Our interpretation of participants’ behaviour is that they wanted to address their gaze at the interlocutor, but finding the panel in front of them made them shift their pupil direction to the top-right or left in some cases. Gaze to “Addressee” was also easily discernible from “Other,” which was again annotated when participants looked at the experimenter by clearly turning their heads to the side. Note that gaze was observed in video recordings with a camera in front of each speaker and not by means of eye-tracking glasses, which, despite providing very good visibility of the eyes, do not enable maximum precision for registering smaller details such as pupil movements. Consequently, these broadly defined categories matched our means and scope.

We did not include the category “Other” in our analysis for either of the two conditions because it is not relevant to our research question.

We did not establish a minimum duration threshold for eye gaze annotation; instead, we annotated all perceivable changes from one of the mentioned gaze directions to another. However, to make a comparison to studies which annotated gaze based on a minimum possible duration threshold (Beattie, 1979; Jokinen et al., 2009; Br ne et al., 2017; Zima et al., 2019; Bavelas et al., 2002, among others), the shortest duration value found in our corpus is 0.069 s, which is below most thresholds previously used (see Degutye and Astell, 2021 for a comprehensive list).

A portion of 20% of the data was annotated independently by a fellow linguist trained in eye gaze annotation on video recordings.

The inter-annotator reliability of gaze direction annotations was assessed using Staccato (Lücking et al., 2011), which is directly available in ELAN and provides a reliable measure of the temporal overlap between annotations based on Thomann's technique (2001). Staccato provided a mean degree of overlap of 70%, which can be interpreted as substantial reliability. To ensure a homogenous interpretation of the category labels for gaze direction, the two annotators discussed labels in the case of disagreement until a common decision was reached. The output annotations performed by the first author are retained in the analysis. An example of all annotation tiers in ELAN is shown in Figure 1.

4.2 Data treatment for the analysis

Annotation of eye gaze was performed for rounds 1, 2, 9, 10, 21, and 22, for each dyad in both conditions. The selected game rounds correspond to the beginning (rounds 1–2), medial part (rounds 9–19), and final part (rounds 21–22) of the whole game session. This selection was made to control for any effect of (non-)familiarity with the task on participants' communicative style. The total number of rounds included in the present multimodal analysis corresponds to 1.47 h of conversation.

To answer our research questions (RQs), we analysed the co-occurrence of vocal feedback signals with gaze directions of both the feedback producer, or “secondary speaker,” and feedback receiver, or “primary speaker” (Section 4.1). To do so, for each participant's vocal feedback signal, we extracted the overlapping gaze direction intervals by both the feedback producer and feedback receiver (RQs 1.a and 1.b, Section 4.1.1) and measured the duration of these intervals up to the moment in which the vocal feedback signal is uttered (RQ 1.c, Section 4.1.2). This duration is operationalised as the time window from the onset of the overlapping gaze direction interval to the onset of the vocal feedback signal. We also extracted the gaze direction intervals overlapping the offset of each feedback signal to establish whether the production of vocal feedback would prompt a shift in gaze direction in any of the two participants (RQ 1.d, Section 4.1.3).

Moreover, we measured the overall amount and duration (in seconds) of gaze direction intervals (RQ 2.a, Section 4.2), as well as vocal feedback rate, operationalised as the number of feedback occurrences per minute of dialogue (RQ 2.b, Section 4.3).

The results of all analyses are presented across visibility conditions (RQ 2.c throughout the result sections for conciseness), i.e., eye-contact (EC) and non-eye-contact (NEC). The data that support the findings of this study, as well as the code used to perform the analysis, are openly available in the accompanying repository at <https://osf.io/3dqzk/>.

5 Results

Although a comparison between the EC and NEC conditions belongs to the second RQ, for reasons of conciseness, we present results relative to the EC and NEC conditions side by side for each measurement.

5.1 Gaze behaviour by the feedback producer and receiver

5.1.1 Occurrences of gaze direction intervals at feedback production

Figure 4 shows the co-occurrences of gaze direction intervals and vocal feedback productions in two graphs: the perspective of the feedback producer (“secondary speaker,” left panel) and the one of the feedback receiver (“primary speaker,” right panel).

In the EC condition, the feedback producers utter most of their turn-regulating vocal feedback while looking at the task (gaze to “Addressee” occurs only in 12% of the cases for PR signals and 9% of the cases for IS signals), contrary to our expectations where we predicted a greater use of gaze to “Addressee” as a way to signal availability as listener.

The results in the EC condition for the feedback receiver reveal a different picture. PR signals occur in 64% of the cases when the primary speaker is looking at the task. This shows that gaze is less frequently used as a request for feedback signals (36%) and confirms our expectations that PR feedback signals are mainly uttered when the primary speaker is not looking at the secondary speaker, not signalling availability for ceding the floor. IS signals occur in equal proportions for both gaze directions: 49% of the cases in which the primary speaker is looking at the task and 51% of the cases in which the primary speaker is looking at the secondary speaker, whereby we expected the latter case to be prevalent if gaze is mainly used to signal turn management. However, we can still observe that being looked at prompts more IS than PR feedback signals, which is in line with our expectations.

In the NEC condition, gaze to “Task” is predominant during feedback for both feedback producer and receiver. In absolute numbers, the low percentages of gaze to “Addressee” correspond to one single instance for PR (the 0% stands for 0.267%) by the feedback producer, and six tokens for PR and three for IS by the feedback receiver. This is also in line with our expectation that the absence of visibility reduces the use of gaze as the signalling function is not available for communicative purposes.

5.1.2 Gaze duration before the feedback utterance

Figure 5 shows the distributions of duration for each gaze direction interval, calculated from the beginning of the interval up to the moment when turn-regulating vocal feedback signals are produced. In other words, this measure indicates how long the feedback producer or receiver maintains the same gaze direction when vocal feedback signals are produced.

Different trends can be observed for the participants in the EC condition (top panels). In the case of the feedback producer, both gaze to “Task” and “Addressee” behave similarly for both feedback pragmatic functions: For PR, most data are in a range of 0–5 s; for IS, most data are located between 0 and 2.5 s. This suggests that when the feedback producer has the intention to take the floor with a vocal feedback signal (IS, turn-initiating feedback), they do so quickly after a change in gaze direction (either to “Task” or “Addressee”).

In the case of the feedback receiver, the distribution for gaze duration to “Addressee” is concentrated around shorter values (0–2.5 s) than gaze duration to “Task,” for which the distribution is wider and spreads across longer duration values (0–5 s). This is true

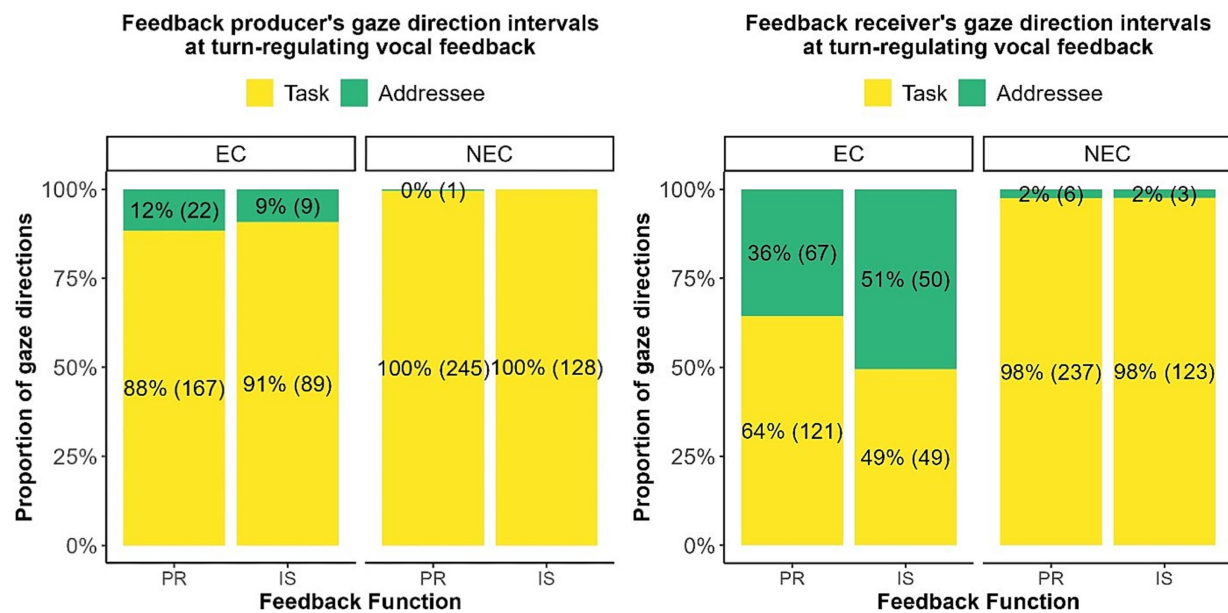


FIGURE 4

Occurrences of gaze direction intervals by the feedback producer and receiver at the moment in which vocal feedback signals are uttered. The proportions of gaze direction intervals are displayed as percentages on the y-axis. The two gaze direction types are colour-coded as in the legend: gaze to task (Task) in yellow and gaze towards addressee (Addressee) in green. The exact proportions and counts of occurrences are printed on the bar plots (the latter in parentheses) for the two turn-regulating functions of the vocal feedback signals, that is, Passive Reciprocity (PR) and Incipient Speakership (IS), displayed on the x-axis. The results are shown for both visibility conditions, that is, eye contact (EC) and non-eye contact (NEC), as indicated on top of the bars.

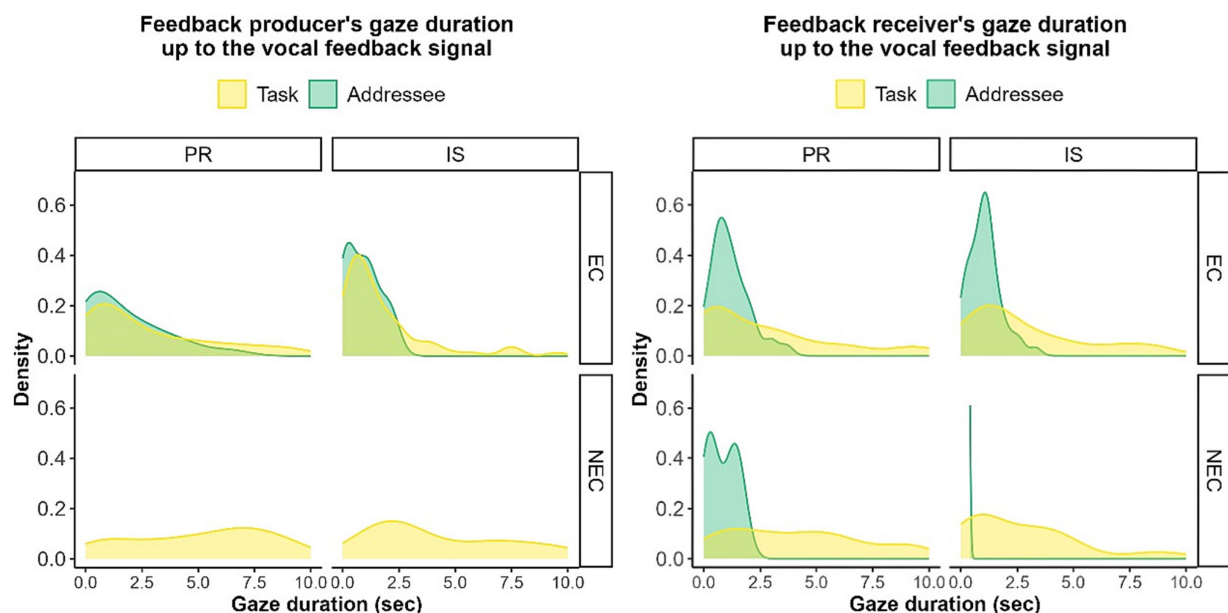


FIGURE 5

Duration of directed gaze to Task (yellow) and Addressee (green) by the feedback producer (left panel) and receiver (right panel) up to the moment vocal feedback signals are uttered. Density values are displayed on the y-axis, and duration of gaze direction intervals (in seconds) is shown on the x-axis. Distributions are shown for the two turn-regulating functions of vocal feedback signals, that is, Passive Reciprocity (PR) and Incipient Speakership (IS). The results are shown for both visibility conditions, eye contact (EC) on top, and non-eye-contact (NEC) on the bottom; note that a limit to the x-axis has been established for improved visualisation as the values show a flat distribution up to 120 s.

for both PR and IS feedback signals, indicating proportionally longer gaze duration to “Task” than to “Addressee” up to the onset of the vocal feedback production.

In the NEC condition, the few instances of gaze to “Addressee” by the feedback receiver, especially for PR, show similar values to those in the EC condition, which might be explained by the evidence

reported in previous studies (see Section 1.4) that speakers perceive the head orientation of their interlocutor even in the absence of visibility. This is, however, not the case for the feedback producer, where there is no gaze to “Addressee.”

5.1.3 Gaze shift after feedback production

Figure 6 shows the occurrences of gaze shifts *after* the production of vocal feedback. A similar trend across interlocutors and visibility conditions can be noticed: The production of turn-regulating vocal feedback does not lead to a subsequent shift in gaze direction.

A slight difference can be observed between the feedback producer and recipient in the EC condition. For the feedback producer, there is a shift in gaze direction in only 9 and 8% of the cases for PR and IS, respectively. For PR, these few cases mostly imply shifting the gaze from the task towards the feedback receiver (12 cases, possibly confirming the availability as a listener). For IS, there is an equal number of shifts towards the task and the feedback receiver (four cases each). A shift in gaze direction after the production of vocal feedback by the receiver occurs in 20% of PR and 15% of IS vocal feedback signals, which is more than what we observe for the feedback producer. For PR, the gaze shift is almost exclusively from addressee to the task (32 cases). For IS, there is the same prevalence of shifts from “Addressee” towards “Task” (10 cases and 5 from task to addressee). In other words, the few cases in which vocal feedback is associated with a shift in gaze direction by the primary speaker occur when the primary speaker looks at the addressee before receiving feedback, and back at the task after ensuring that the listener is following the conversation.

However, overall, there are very few data points of gaze shift after a vocal feedback signal across the two visibility conditions. Therefore,

it appears that vocal feedback signals rarely occur with changes in the gaze direction of the interlocutors.

5.2 Overall gaze direction intervals and their duration

We found 1,649 gaze direction intervals in the eye-contact condition (EC; 80% of the total) and 354 gaze direction intervals in the non-eye-contact condition (NEC; 20% of the total), showing that, in line with our expectations, participants change the direction of their gaze far less often when eye contact is inhibited.

However, the relative proportions of gaze direction intervals directed to “Task” or “Addressee” displayed in Figure 7 (y-axis) do not differ greatly across conditions. In the EC condition, there is an almost equal proportion of gaze occurrences to “Task” and “Addressee” (50.5 and 49.5% respectively). In the NEC condition, where the interlocutor is behind an opaque panel, the proportion of gaze to “Task” increases only slightly (60.7%) as compared to the EC condition. On the other hand, the proportion of gaze to “Addressee” (39.3%), although lower than in the EC condition, indicates that participants gaze away from the task quite often, looking at the panel, that is, in the direction of the interlocutor’s voice, namely, where the addressee is known to be sitting.

Mean gaze duration (Figure 8) shows that the time spent looking at the “Task” is longer than that spent looking towards the “Addressee” in both conditions (4.75 s in EC and 18.2 s in NEC). However, the time spent looking at the “Task” shows high variability and can potentially be longer, especially when eye contact is inhibited (see large error bars especially for NEC). There is less variability in the time spent looking towards the “Addressee,”

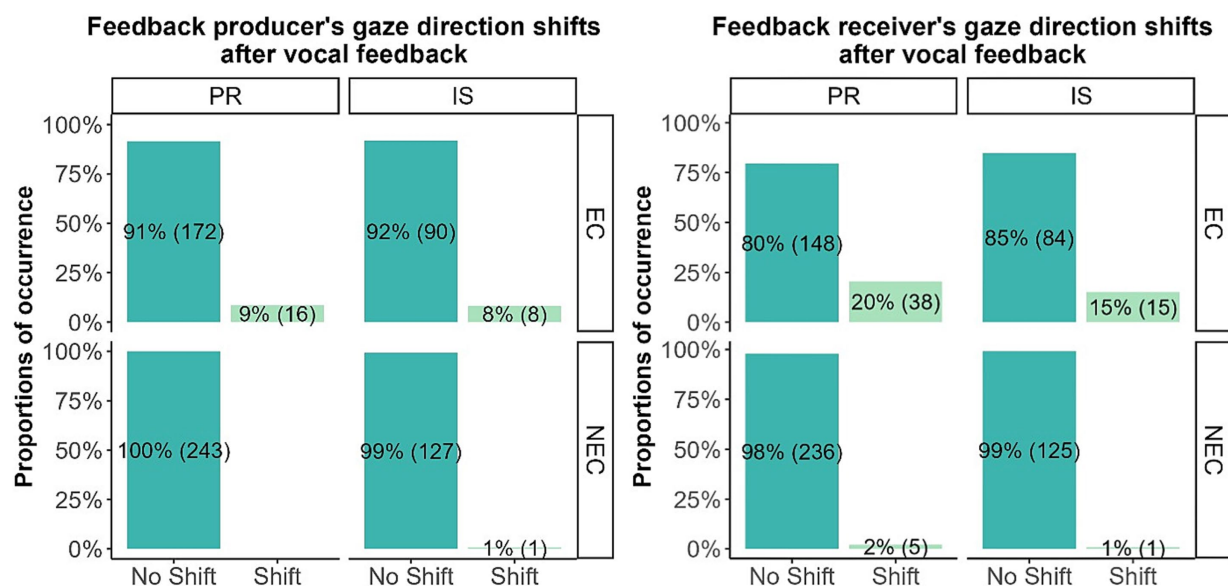
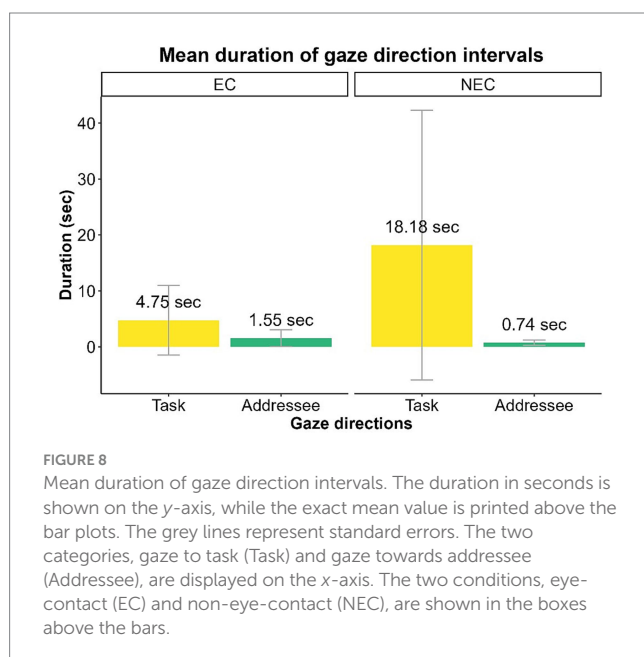
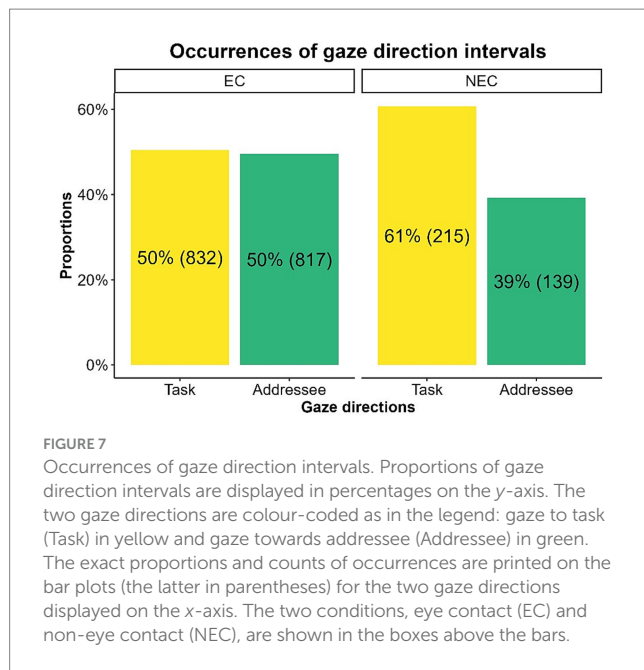


FIGURE 6

Shifts in gaze direction by the feedback producer and receiver after the production of a vocal feedback signal. The proportions of gaze direction shifts are displayed as percentages on the y-axis. Exact proportions and counts of occurrences are printed on the bar plots (the latter in parenthesis) for the two categories “No Shift” and “Shift” in gaze direction, displayed on the x-axis and colour-coded. Occurrences are shown for the two turn-regulating functions of vocal feedback signals, that is, Passive Reciprocity (PR) and Incipient Speakership (IS), indicated on the top, and across visibility conditions, i.e., eye-contact (EC) and non-eye-contact (NEC), indicated on the right.

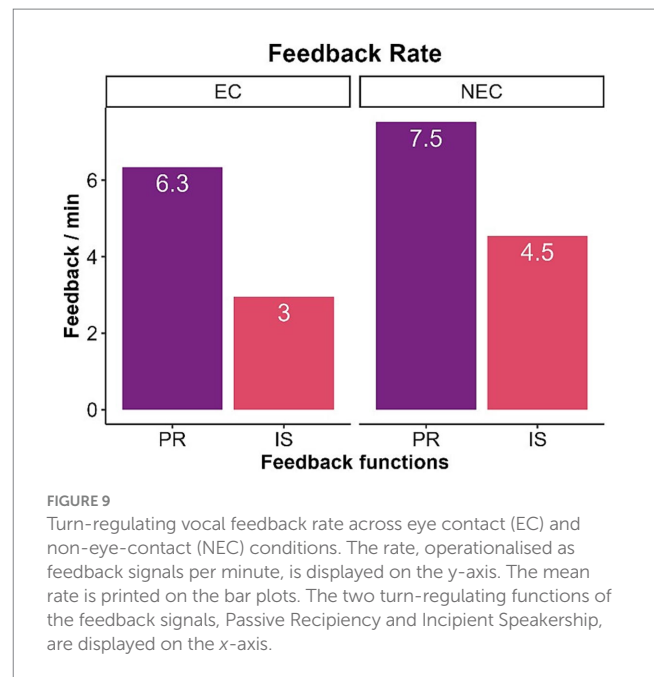


with a mean duration of 1.55 s in the EC condition and 0.74 s in the NEC condition, especially in the NEC condition (small error bars).

In line with our expectations, these two metrics taken together show that the lower number of gaze direction intervals in the NEC condition could be explained by the fact that participants gaze for longer periods at the task than in the EC condition, leaving little room for switching gaze direction.

5.3 Overall vocal feedback rate

Finally, [Figure 9](#) shows the rate of vocal feedback production across pragmatic functions and visibility conditions. These values



indicate that the vocal feedback rate is higher in the NEC than in the EC condition, as expected. Across feedback functions, our analysis reveals a proportionally higher rate (almost double) of PR than IS vocal feedback expressions.

6 Discussion

Our first research question (RQ 1 in Section 2) concerned the interplay between gaze and vocal turn-regulating feedback by the feedback producer and receiver in terms of the number of gaze direction intervals, their duration, and the shift of gaze direction after the production of vocal feedback.

Our results show that, in a face-to-face task-based interaction, the feedback producer (the secondary speaker, RQ 1a) almost exclusively looks at the task while producing vocal feedback, independent of the intention to take the floor. This result suggests that feedback producers might rely on cues other than eye gaze to signal the floor management function of vocal feedback. For example, they may use prosody (in particular, intonational cues), which has been found to vary according to the turn-regulating function and type of feedback ([Savino, 2010, 2011, 2012](#); [Savino and Refice, 2013](#); [Sbranna et al., 2022](#); [Sbranna et al., 2024](#)).

The results for gaze behaviour of the feedback receiver (the primary speaker, RQ 1b) are different, with a higher percentage of gaze towards the Addressee for both feedback functions, but especially for IS (when the secondary speaker intends to take the floor). The latter case is in line with the literature, in which the primary speaker is said to gaze at the interlocutor when signalling a speaker change. Since interactions included a visual task, it is understandable that we find a higher level of gaze to task than predicted by previous studies on gaze and turn alternation, which were mostly based on free conversations and reported mutual gaze in correspondence with feedback. Nevertheless, our data show a trend in the expected direction for the feedback receiver: gaze to task, showing unavailability

for a turn transition, mostly prompts continuers (PR), whereas gaze to addressee prompts a high percentage of turn-initial feedback (IS). Comparing these results with those obtained by Spaniol et al. (2023), we can confirm the reported prevalence of averted gaze together with vocal feedback by the feedback producer. However, in our dataset, we also find that the feedback receiver uses a greater amount of gaze directed towards the interlocutor when the interlocutor is producing vocal feedback. In relative terms across functions, gaze is produced less with PR and more with IS, showing that the signalling function of gaze is only marginally used to elicit feedback and is more often used to regulate turns. This more fine-grained description of gaze, going beyond the binary distinction between ‘mutual’ and ‘averted’ gaze is a possible explanation for the difference between our results and previous ones.

We also investigated the duration of the gaze direction intervals up to the moment when vocal feedback signals are uttered (RQ 1c) to explore the temporal details (as in Kendrick et al., 2023 and Ondáš et al., 2023) of the time window of gaze under which the vocal feedback is uttered. We observe a similar range of duration values as Ondáš et al. (2023) for the feedback receiver, who obtains a quick vocal response after initiating gaze, independent of the feedback turn-regulating function. This suggests that looking at the secondary speaker may be primarily interpreted by the latter as a feedback request. The findings on these temporal details can also be related to the concept of an advantageous mutual gaze window for feedback insertion proposed in previous studies (Bavelas et al., 2012; Bavelas et al., 2002). From our dataset, it emerges that an advantageous gaze window is not necessarily mutual—since the occurrences of gaze to “Addressee” while uttering feedback, i.e., by the secondary speaker, are much lower than those of the primary speaker—but can be *unilateral* as even the sole activation of gaze by the primary speakers elicits a quick vocal response. Instead, for the feedback producer, the data suggest a tendency for shorter gaze duration (either to “Task” or “Addressee”) before turn-initiating feedback signals, i.e., the feedback producer quickly takes the floor with a vocal feedback signal after a change in their gaze direction. This result contrasts with the findings of Kendrick et al. (2023), who claimed that the direction of speaker gaze does not affect the speed of general turn transitions. The secondary speaker might consider moments of perturbations of the previous state of balance in the conversation, that is, in proximity to a change in activation/deactivation of visual cues from their side, as better suited for a change in speakership (see Complex Dynamic System Theory, Cameron and Larsen-Freeman, 2007).

Finally, we find no evidence that the production of turn-regulating vocal feedback causes a shift in gaze direction by either the feedback producer or receiver in most cases (RQ 1d), as found by Bavelas et al. (2002). This result might also be interpreted in the light of our task-based setting, in which visual attention to the task materials was necessary to complete the task: The overall longer gaze time to task during the dialogues increases the probability that vocal feedback (at both their onset and offset) corresponds to these gaze intervals. However, in the very few cases in which a gaze shift occurs, it is prevalent in the feedback receiver after PR signals, which is in line with the findings of Bavelas et al. (2002). Similar to their findings, these shifts are from gaze to addressee towards the task, showing that the visual contact established by the primary speaker might indeed have the function of eliciting vocal feedback as it is concluded right after its utterance.

Our second research question (RQ 2 as in Section 2) concerned the effect of non-visibility between participants on their use of gaze and vocal feedback.

One striking remark on overall gaze behaviour is that, despite the presence of a visual competitor in this experimental setting, participants still show high relative proportions of occurrences of gaze directed towards the interlocutor, independent of whether eye contact was inhibited or not (in line with Edlund et al., 2012; Kato et al., 2010; Nakano et al., 2008 about the ability of speakers to identify interlocutors’ head direction in darkness, based on acoustic cues). The greatest difference in overall gaze behaviour across visibility conditions was related to the length of these gaze direction intervals, with participants spending more time looking at the task when visual contact was impeded (RQ 2a). This result is in line with previous studies reporting that non-visibility between interlocutors in a conversation causes a drastic drop in the search for visual contact (Argyle et al., 1973; Renklint et al., 2012). This might be explained by the fact that in the EC condition, gaze fulfils a communicative signalling function in addition to its primary sensing function, whereas in the NEC condition only the sensing function is possible—we impeded eye contact, but participants could still sense their environment—reducing the occasions in which switching gaze direction is necessary. This result, while seemingly trivial at first, opens up key future questions as to which functions eye gaze fulfil and how these functions are distributed when eye contact is impeded. Since we found that the difference in the proportion of occurrences of gaze to addressee is only 10% across the two conditions, the question of which speech events relate to these short but still represented occurrences of gazes to addressee in the absence of visual contact remains open as no correspondence with turn-regulating feedback (2b) was found. One speculation we propose is that eye gaze directed towards the sound source in non-visibility conditions may be activated at moments of heightened attention, serving a focusing function on the source of information in response to increased cognitive demand.

Vocal feedback rate shows that the unavailability of the visual channel boosts the production of vocal feedback signals (RQ 2c), which is in line with previous findings of increased use of backchannels and vocal resources in general in the absence of eye contact (e.g., Cook and Lalljee, 1972; Boyle et al., 1994; Neiberg and Gustafson, 2011). We also found an overall prevalence of PR over IS vocal feedback, which has also been reported in previous task-based studies on vocal feedback in Tangram game-based dyadic conversations in German (Spaniol et al., 2023) and Map Tasks (Anderson et al., 1991) in Italian and German (Savino, 2010, 2011, 2012; Savino and Refice, 2013; Sbranna et al., 2024; Wehrle, 2023; Sbranna et al., 2022). Thus, this phenomenon might be related to a collaborative-task-based setting, in which speakers tend to alternately lead the conversation, leaving the interlocutor the role of acknowledging the other’s speech to ensure that the information necessary to complete the task has been successfully received, that is, to update the common ground/knowledge (Clark and Schaefer, 1989).

7 Conclusion

We conducted a small-sample study on the interplay between gaze and turn-regulating vocal feedback, with a 2-fold goal. Using

video-recorded Tangram game conversations by six dyads of Italian speakers in eye-contact and non-eye-contact conditions, we analysed (1) the relationship between turn-regulating vocal feedback—acknowledgements with the functions of “Incipient Speakership” (turn-initiating) and “Passive Reciprocity” (continuers)—and eye gaze directions (to the Task or the Addressee) and (2) how far this relation, as well as overall gaze and feedback behaviour, change when eye contact is impeded. In the following paragraphs, we address the limitations and innovations of the study.

First, this analysis was based on a relatively limited number of participants. Increasing the sample size would allow for robust statistical testing of these preliminary findings, and therefore, for a certain degree of generalisability of the phenomena reported here. Most previous studies on gaze and turn alternation were based on free conversation, whereas our task-based design was meant to match a comparable study design on gaze and turn-regulating feedback. A task-based setting should not be regarded as less generalisable to real-world situations than free conversations, since many real-world dialogues involve a visual competitor, that is, an object in the immediate environment that is also the topic of conversation. These conversational contexts are even more challenging for interlocutors as they have to compromise between looking at the object and using gaze and speech to effectively manage the conversation. However, to guarantee full comparability with previous studies and partial out any variation from previous studies due to different designs, a replication of this study with free conversations, including an analysis of eye gaze at turn transition for comparison, would be beneficial. This would shed light on the differences and similarities between eye gaze during turn alternation and turn-regulating vocal feedback. Moreover, such an analysis should account for the content of the sentences involved in the turn alternations (e.g., adjacency pairs) as gaze behaviour may vary depending on pragmatic aspects.

Second, we did not use eye tracking technology, meaning that our annotations were made from the observer's perspective, which may differ from the participant's perspective elicited in an eye tracking study. For our research goal, this did not represent a problem as participants were video recorded frontally and eye movements were fully detectable. However, the complementary use of eye-tracking techniques in the future, apart from allowing automated analysis of larger datasets, could provide a higher level of detail, especially useful for further investigating the more variable nature of the addressee-oriented gaze we found in the absence of eye contact.

Finally, our investigation is limited to the relationship between eye gaze and turn-regulating vocal feedback, but expanding the current study to visual types of feedback such as head nods and including prosodic aspects of feedback would provide a more complete picture.

Despite these shortcomings, we provided insightful preliminary findings regarding the under-researched relationship between eye gaze and turn-regulating feedback. As compared to previous designs based on a binary distinction of ‘mutual’ vs. ‘averted’ gaze, we proposed a tripartite distinction including unilateral directed gaze, and analysing gaze by the feedback producer and receiver, which provided novel insights. In particular, the finding that the feedback receiver (the primary speaker) uses gaze more often than the feedback producer

(the secondary speaker) suggests that *unilateral* gaze (i.e., when gaze by one participant is not reciprocated by the interlocutor) is also used as a successful and efficient feedback request. Furthermore, we provided new findings about the interplay between vocal feedback and gaze behaviour in a task-based context in the absence of eye contact, which, to our knowledge, had not been investigated before. We found that gaze is directed towards the interlocutor even in the absence of eye contact, but not for turn-regulating vocal feedback, opening new research perspectives. Future research examining unilateral and mutual gaze will help confirm and expand these findings to provide a more complete picture of these interactional phenomena.

Data availability statement

The data table and analysis code presented in this study can be found at <https://osf.io/3dqzk/>.

Ethics statement

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

SS: Conceptualization, Data curation, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing. MS: Methodology, Writing – original draft, Writing – review & editing, Data curation, Investigation, Resources. FB: Methodology, Writing – original draft, Writing – review & editing. MG: Methodology, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was funded by the Cluster Development Program Language Challenges, a funding line within the Excellent Research Support Program of the University of Cologne, and the German Research Foundation (DFG), grant number 281511265, SFB 1252 Prominence in Language.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

References

- Alibali, M. W., Heath, D. C., and Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: some gestures are meant to be seen. *J. Mem. Lang.* 44, 169–188. doi: 10.1006/jmla.2000.2752
- Amador-Moreno, C. P., McCarthy, M., and O'Keeffe, A. (2013). "Can English provide a framework for Spanish response tokens?" in *Yearbook of Corpus Linguistics and Pragmatics 2013: New Domains and Methodologies* (Dordrecht: Springer Netherlands), 175–201.
- Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., et al. (1991). The HCRC map task corpus. *Lang. Speech* 34, 351–366. doi: 10.1177/002383099103400404
- Argyle, M., and Cook, M. (1976). *Gaze and Mutual Gaze*. Cambridge: Cambridge University Press.
- Argyle, M., and Dean, J. (1965). Eye-contact, distance and affiliation. *Sociometry* 28:289. doi: 10.2307/2786027
- Argyle, M., Ingham, R., Alkema, F., and McCallin, M. (1973). The different functions of gaze. *Semiotica* 7, 19–32. doi: 10.1515/semi.1973.7.1.19
- Argyle, M., Lalljee, M., and Cook, M. (1968). The effects of visibility on interaction in a dyad. *Human relations* 21, 3–17.
- Auer, P. (2018). "Gaze, addressee selection and turn-taking in three-party interaction," in *Eye-tracking in interaction: Studies on the role of eye gaze in dialogue*. eds. G. Brône and B. Oben (Amsterdam: John Benjamins), 197–232.
- Bavelas, J. B., Chovil, N., Lawrie, D. A., and Wade, A. (1992). Interactive gestures. *Discourse Process.* 15, 469–489. doi: 10.1080/01638539209544823
- Bavelas, J. B., Coates, L., and Johnson, T. (2002). Listener responses as a collaborative process: the role of gaze. *J. Commun.* 52, 566–580. doi: 10.1111/jcom.2002.52.issue-3
- Bavelas, J. B., De Jong, P., Korman, H., and Jordan, S. S. (2012). Beyond back-channels: A three-step model of grounding in face-to-face dialogue. In *Proceedings of Interdisciplinary Workshop on Feedback Behaviors in Dialog*, 5–6.
- Bavelas, J. B., Gerwing, J., Sutton, C., and Prevost, D. (2008). Gesturing on the telephone: independent effects of dialogue and visibility. *J. Mem. Lang.* 58, 495–520. doi: 10.1016/j.jml.2007.02.004
- Beattie, G. W. (1978). Floor apportionment and gaze in conversational dyads. *British J. Soc. Clin. Psychol.* 17, 7–15. doi: 10.1111/j.2044-8260.1978.tb00889.x
- Beattie, G. W. (1979). Contextual constraints on the floor-apportionment function of speaker-gaze in dyadic conversations. *Br. J. Soc. Clin. Psychol.* 18, 391–392. doi: 10.1111/j.2044-8260.1979.tb00909.x
- Bertrand, R., Ferré, G., Blache, P., Espesser, R., and Rauzy, S. (2007). "Backchannels revisited from a multimodal perspective," in *Proceedings of the ISCA Workshop "Auditory Audio-Visual Processing"*. Available at: https://www.isca-archive.org/avsp_2007/bertrand07_avsp.pdf.
- Bissonnette, V. L. (1993). Interdependence in dyadic gazing [doctoral dissertation, the University of Texas at Arlington]. Ann Arbor: ProQuest Dissertations Publishing.
- Blythe, J., Gardner, R., Mushin, I., and Stirling, L. (2018). Tools of engagement: selecting a next speaker in Australian aboriginal multiparty conversations. *Res. Lang. Soc. Interact.* 51, 145–170. doi: 10.1080/08351813.2018.1449441
- Boersma, P., and Weenink, D. (2001). Praat, a system for doing phonetics by computer. *Glott International* 5, 341–345.
- Borràs-Comes, J., Kaland, C., Prieto, P., and Swerts, M. (2014). Audiovisual correlates of interrogativity: a comparative analysis of Catalan and Dutch. *J. Nonverbal Behav.* 38, 53–66. doi: 10.1007/s10919-013-0162-0
- Boyle, E., Anderson, A., and Newlands, A. (1994). The effects of visibility on dialogue and performance in a cooperative problem-solving task. *Lang. Speech* 37:1.
- Brône, G., Oben, B., Jehoul, A., Vranjes, J., and Feyaerts, K. (2017). Eye gaze and viewpoint in multimodal interaction management. *Cog. Linguis.* 28, 449–483. doi: 10.1515/cog-2016-0119
- Cameron, L., and Larsen-Freeman, D. (2007). Complex systems and applied linguistics. *Int. J. Appl. Linguist.* 17, 226–240. doi: 10.1111/j.1473-4192.2007.00148.x
- Cañigueral, R., and Hamilton, A. F. C. (2019). The role of eye gaze during natural social interactions in typical and autistic people. *Front. Psychol.* 10:560. doi: 10.3389/fpsyg.2019.00560
- Clancy, P. M., Thompson, S. A., Suzuki, R., and Tao, H. (1996). The conversational use of reactive tokens in English, Japanese, and Mandarin. *J. Pragmat.* 26, 355–387. doi: 10.1016/0378-2166(95)00036-4
- Clark, H. H., and Schaefer, E. F. (1989). Contributing to discourse. *Cogn. Sci.* 13, 259–294. doi: 10.1207/s15516709cog1302_7
- Cook, M., and Lalljee, M. G. (1972). Verbal substitutes for visual signals in interaction. *Semiotica* 6, 212–221. doi: 10.1515/semi.1972.6.3.212
- Cummins, F. (2012). Gaze and blinking in dyadic conversation: a study in coordinated behaviour among individuals. *Lang. Cog. Proc.* 27, 1525–1549. doi: 10.1080/01690965.2011.615220
- Degutye, Z., and Astell, A. (2021). The role of eye gaze in regulating turn taking in conversations: a systematized review of methods and findings. *Front. Psychol.* 12:616471. doi: 10.3389/fpsyg.2021.616471
- Drummond, K., and Hopper, R. (1993). Back channels revisited: acknowledgment tokens and speakership incipency. *Res. Lang. Soc. Interact.* 26, 157–177. doi: 10.1207/s15327973rlsi2602_3
- Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *J. Personal. Soc. Psychol.* 23, 283–292. doi: 10.1037/h0033031
- Duncan, S., and Fiske, D. W. (1977). *Face-to-face interaction: Research, methods, and theory*. London: Routledge.
- Eberhard, K. M., and Nicholson, H. (2010). Coordination of understanding in face-to-face narrative dialogue. In *Proceedings of the Annual Meeting of the Cognitive Science Society*.
- Edlund, J., Heldner, M., and Gustafson, J. (2012). "On the effect of the acoustic environment on the accuracy of perception of speaker orientation from auditory cues alone," in *The Proceedings of 13th Annual Conference of the International Speech Communication Association*. pp. 1482–1485.
- Egbert, M. M. (1996). Context-sensitivity in conversation: Eye gaze and the German repair initiator *bitte*? *Language in Society* 25, 587–612.
- Felleg, A. M. (1995). Patterns and functions of minimal response. *Amer. Speech Int. J. Educ. Best Prac.* 70, 186–199. doi: 10.2307/455815
- Ferré, G., and Renaudier, S. (2017). Unimodal and bimodal backchannels in conversational English. In *Proceedings of SemDial*, Saarbrücken, Germany, 27–37.
- Foulsham, T., Cheng, J. T., Tracy, J. L., Henrich, J., and Kingstone, A. (2010). Gaze allocation in a dynamic situation: effects of social status and speaking. *Cognition* 117, 319–331. doi: 10.1016/j.cognition.2010.09.003
- Fries, C. C. (1952). *The structure of English*. London: Longmans, Green; Company.
- Gardner, R. (2001). *When listeners talk: Response tokens and listener stance*. Amsterdam: John Benjamins Publishing Company.
- Gobel, M. S., Kim, H. S., and Richardson, D. C. (2015). The dual function of social gaze. *Cognition* 136, 359–364. doi: 10.1016/j.cognition.2014.11.040
- Goodwin, C. (1980). Restarts, pauses, and the achievement of a state of mutual gaze at turn-beginning. *Sociol. Inquiry* 5, 272–302. doi: 10.1111/j.1475-682X.1980.tb00023.x
- Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*. London: Academic Press.
- Harrigan, J. A., and Steffen, J. J. (1983). Gaze as a turn-exchange signal in group conversations. *Br. J. Soc. Psychol.* 22:167–168. doi: 10.1111/j.2044-8309.1983.tb00578.x
- Hasegawa, Y. (2014). *Japanese: A linguistic introduction*. Cambridge: Cambridge University Press.
- Heldner, M., Hjalmarsson, A., and Edlund, J. (2013). Backchannel relevance spaces. In E. L. Asu and P. Lippus (Eds.), *Nordic Prosody: Proceedings of the XIth Conference, Tartu 2012* (pp. 137–146).
- Heritage, J. (1984). "A change-of-state token and aspects of its sequential placement" in *Structures of social action: Studies in conversation analysis*. Cambridge: Cambridge University Press, 299–345.
- Hjalmarsson, A., and Oertel, C. (2012). "Gaze direction as a back-channel inviting cue in dialogue" in *Proceedings of the IVA Workshop on "Realtime conversational virtual agents"*. vol. 9 (Citeseer).
- Ho, S., Foulsham, T., and Kingstone, A. (2015). Speaking and listening with the eyes: gaze signaling during dyadic interactions. *PLoS One* 10, 1–18. doi: 10.1371/journal.pone.0136905
- Hollenstein, T. (2013). *State space grids*. Boston, MA: Springer.
- Hu, L., and Chen, G. (2021). Trajectories of idea emergence in dialogic collaborative problem solving: toward a complex dynamic systems perspective. *Front. Psychol.* 12:735534. doi: 10.3389/fpsyg.2021.735534

- Janz, A. (2022). Navigating common ground using feedback in conversation – A phonetic analysis. Cologne, Germany: University of Cologne.
- Jefferson, G. (1983). Two explorations of the organization of overlapping talk in conversation: Notes 915 on some orderlinesses of overlap onset. *Tilburg Papers in Language and Literature* 28, 1–28.
- Jokinen, K., Furukawa, H., Nishida, M., and Yamamoto, S. (2013). Gaze and turn-taking behavior in casual conversational interactions. *ACM Trans. Interact. Intell. Syst.* 3, 1–30. doi: 10.1145/2499474.2499481
- Jokinen, K., Nishida, M., and Yamamoto, S. (2009). Eye-gaze experiments for conversation monitoring. In Proceedings of the 3rd International Universal Communication Symposium (pp. 303–308)
- Jurafsky, D., Shriberg, E., Fox, B., and Curl, T. (1998). “Lexical, prosodic, and syntactic cues for dialog acts,” in *Proceedings of the ACL Workshop “Discourse Relations and Discourse Markers”*. 114–120.
- Kalma, A. (1992). Gazing in triads: a powerful signal in floor apportionment. *Br. J. Soc. Psychol.* 31, 21–39. doi: 10.1111/j.2044-8309.1992.tb00953.x
- Kato, H., Takemoto, H., Nishimura, R., and Mokhtari, P. (2010). Spatial acoustic cues for the auditory perception of speaker's facing direction. In Proc. of 20th International Congress on Acoustics, ICA 2010. Sydney, Australia
- Kawahara, T., Iwatate, T., and Takanashi, K. (2012). Prediction of turn-taking by combining prosodic and eye-gaze information in poster conversations. In R. Sproat (Ed.), *Proceedings of Interspeech* (pp. 727–730).
- Kendon, A. (1967). Some functions of gaze direction in social interaction. *Acta Psychol.* 26, 22–63. doi: 10.1016/0001-6918(67)90005-4
- Kendon, A. (1990). Conducting interaction: Patterns of behavior in focused encounters, vol. 7: CUP Archive.
- Kendrick, K. H., Holler, J., and Levinson, S. C. (2023). Turn-taking in human face-to-face interaction is multimodal: gaze direction and manual gestures aid the coordination of turn transitions. *Philos. Trans. Royal Soc B* 378:20210473. doi: 10.1098/rstb.2021.0473
- Kraut, R. E., Lewis, S. H., and Swezey, L. W. (1982). Listener responsiveness and the coordination of conversation. *J. Personal. Soc. Psychol.* 43, 718–731. doi: 10.1037/0022-3514.43.4.718
- Lambertz, K. (2011). Back-channelling: the use of yeah and mm to portray engaged listenership. Griffith Working Papers in Pragmatics and Intercultural Communication, 4, 11–18.
- Lerner, G. H. (2003). Selecting next speaker: the context-sensitive operation of a context-free organization. *Lang. Soc.* 32, 177–201. doi: 10.1017/S00474045032202X
- Levinson, S. C., and Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Front. Psychol.* 6:731. doi: 10.3389/fpsyg.2015.00731
- Lücking, A., Ptock, S., and Bergmann, K. (2011). “Assessing agreement on segmentations by means of staccato, the segmentation agreement calculator according to Thomann,” in *International Gesture Workshop* (Berlin, Heidelberg: Springer Berlin Heidelberg), 129–138.
- McNeill, D. (2005). Gesture, gaze, and ground. In machine learning for multimodal interaction: second international workshop. Berlin Heidelberg: Springer.
- Myszka, T. J. (1975). Situational and intrapersonal determinants of eye contact, direction of gaze aversion, smiling and other non-verbal behaviors during an interview [doctoral dissertation, University of Windsor]. Electronic Thesis and Dissertations University of Windsor. Available online at: <https://scholar.uwindsor.ca/etd/3477>
- Nakano, A. Y., Yamamoto, K., and Nakagawa, S. (2008). Auditory perception of speaker's position, distance and facing angle in a real enclosed environment. In Proc. of Autumn Meeting of Acoustic Society of Japan (pp. 525–526).
- Neiberg, D., and Gustafson, J. (2011). Predicting speaker changes and listener responses with and without eye-contact. *Proc. Interspeech* 1565–1568. doi: 10.21437/Interspeech.2011-471
- Novick, D., Hansen, B., and Ward, K. (1996). Coordinating turn-taking with gaze. In Proceedings of Fourth International Conference on Spoken Language Processing (ICSLP-96) (pp. 1888–1891)
- Oertel, C., Włodarczyk, M., Edlund, J., Wagner, P., and Gustafson, J. (2012). Gaze patterns in turn-taking. In R. Sproat (Ed.), *Proceedings of Interspeech* 2012 (pp. 2246–2249)
- Ondáš, S., Kiktová, E., Pleva, M., and Juhár, J. (2023). Analysis of backchannel inviting cues in dyadic speech communication. *Electronics* 12:3705. doi: 10.3390/electronics12173705
- Paulmann, S., Jessen, S., and Kotz, S. A. (2009). Investigating the multimodal nature of human communication. *J. Psychophysiol.* 23, 63–76. doi: 10.1027/0269-8803.23.2.63
- Poppe, R., Truong, K. P., and Heylen, D. (2011). Backchannels: quantity, type and timing matters. In Intelligent Virtual Agents: 10th International Conference, IVA 2011, Reykjavik, Iceland, September 15–17, 2011. Proceedings 11 (pp. 228–239)
- Rasenberg, M., Pouw, W., Özyürek, A., and Dingemanse, M. (2022). The multimodal nature of communicative efficiency in social interaction. *Sci. Report* 12:19111. doi: 10.1038/s41598-022-22883-w
- Renklint, E., Cardell, F., Dahlbäck, J., Edlund, J., and Heldner, M. (2012). Conversational gaze in light and darkness. *Proc. Fonetik* 2012, 59–60. Available at: <https://www.diva-portal.org/smash/get/diva2:546623/FULLTEXT01>
- Risko, E. F., Richardson, D. C., and Kingstone, A. (2016). Breaking the fourth wall of cognitive science: real-world social attention and the dual function of gaze. *Curr. Dir. Psychol. Sci.* 25, 70–74. doi: 10.1177/0963721415617806
- Rossano, F. (2012). Gaze behaviour in face-to-face interaction [doctoral dissertation, Radboud university, Nijmegen]. Available online at: <https://hdl.handle.net/11858/00-001M-0000-000F-ED23-5>
- Rossano, F., Brown, P., and Levinson, S. C. (2009). Gaze, questioning and culture. *Convers. Analys.* 27, 187–249. doi: 10.1017/CBO9780511635670.008
- Rühlemann, C. (2007). Conversation in context: A corpus-driven approach. London: Longman.
- Rutter, D. R., Stephenson, G. M., Ayling, K., and White, P. A. (1978). The timing of looks in dyadic conversation. *British J. Soc. Clin. Psychol.* 17, 17–21. doi: 10.1111/j.2044-8260.1978.tb00890.x
- Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 696–735. doi: 10.2307/412243
- Savino, M. (2010). “Intonational strategies for backchanneling in Italian Map Task dialogues,” in *Proceedings of the 3rd ISCA Workshop ExLing* 2010.
- Savino, M. (2011). The intonation of backchannels in Italian task-oriented dialogues: Cues to turn-taking dynamics, information status and speaker's attitude. In Proceedings of the 5th Language and Technology Conference: Human Language Technology as a Challenge for Computer Science and Linguistics.
- Savino, M. (2012). The intonation of polar questions in Italian: Where is the rise? *J. Int. Phonetic Assoc.* 42, 23–48.
- Savino, M., Lapertosa, L., and Refice, M. (2018). “Seeing or Not Seeing Your Conversational Partner: The Influence of Interaction Modality on Prosodic Entrainment” in *Speech and Computer*. eds. A. Karpov, O. Jokisch and R. Potapova (Switzerland: Springer).
- Savino, M., and Refice, M. (2013). Acknowledgement or reply? Prosodic features for disambiguating pragmatic functions of the Italian token “sì.” In Proceedings of the 7th Conference on Speech Technology and Human-Computer Dialogue (SpED) (pp. 1–6).
- Sbranna, S., Wehrle, S., and Grice, M. (2022). “The use of Backchannels and other Very Short Utterances by Italian Learners of German” in *La posizione del parlante nell'interazione: atteggiamenti, intenzioni ed emozioni nella comunicazione verbale [The position of the speaker in interaction: attitudes, intentions, and emotions in verbal communication]*. eds. R. Orrico and L. Schettino (Milan: Officinaventuno).
- Sbranna, S., Wehrle, S., and Grice, M. (2024). A multi-dimensional analysis of backchannels in L1 German, L1 Italian and L2 German. *Lang. Interact. Acquis.* 15.2, 242–276. doi: 10.1075/lia.00026.sbr
- Schegloff, E. A. (1982). “Discourse as an interactional achievement: Some uses of ‘uh huh’ and other things that come between sentences” in *Analysing discourse: Text and talk*. ed. D. Tannen (Washington, D.C.: Georgetown University Press), 71–93.
- Shelley, L., and Gonzalez, F. (2013). Back channeling: function of back channeling and L1 effects on back channeling in L2. *Linguist. Portfolios* 2:9. Available at: https://repository.stcloudstate.edu/stcloud_ling/vol2/iss1/9
- Simon, C. (2018). The functions of active listening responses. *Behav. Process.* 157, 47–53. doi: 10.1016/j.beproc.2018.08.013
- Spaniol, M., Janz, A., Wehrle, S., Vogeley, K., and Grice, M. (2023). “Multimodal signalling: the interplay of oral and visual feedback in conversation,” in *Proceedings of the 20th international congress of phonetic sciences*.
- Stahl, G. (2016). Constructing dynamic triangles together: The development of mathematical group cognition. Cambridge: Cambridge University Press.
- Streeck, J. (2014). “Mutual gaze and recognition. Revisiting Kendon's gaze direction in two-person conversation” in *From gesture in conversation to visible action as utterance*. eds. M. Seyfeddinipur and M. Gullberg (Amsterdam: John Benjamins), 35–55.
- The Language Archive. (2023). Nijmegen: Max Planck Institute for Psycholinguistics, ELAN (Version 6.7). Available online at: <https://archive.mpi.nl/tla/elan>
- Truong, K. P., Poppe, R., Kok, I. D., and Heylen, D. (2011). A multimodal analysis of vocal and visual backchannels in spontaneous dialogs. *Proc. Interspeech* 2011, 2973–2976. doi: 10.21437/Interspeech.2011-744
- Wehrle, S. (2023). Conversation and intonation in autism: A multi-dimensional analysis. Berlin: Language Science Press.
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. In Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation.
- Yngve, V. H. (1970). On getting a word in edgewise. *Chicago Linguistics Society* 6, 567–578.
- Zima, E., Weiß, C., and Bröne, G. (2019). Gaze and overlap resolution in triadic interactions. *J. Pragmat.* 140, 49–69. doi: 10.1016/j.pragma.2018.11.019